

# NGGDPP Metadata Preparation Guide (June 2015)

The National Digital Catalog (NDC) administered by the National Geological and Geophysical Data Preservation Program (NGGDPP) includes essential metadata describing samples and collection sites to promote discovery and encourage further exploration of similar items by identifying contacts, providing access instructions, and linking to online resources for additional information about collections and records. Please follow instructions below to prepare metadata describing geoscience materials for submission to the NDC. Additional information is available at <http://datapreservation.usgs.gov>. Please direct questions to [nggdpp@usgs.gov](mailto:nggdpp@usgs.gov).

## 1 – National Digital Catalog (NDC) Metadata Elements

The tables in this section list the thirteen (13) metadata elements accepted by the NDC. Some metadata elements are mandatory and must be included to adequately describe records. Other metadata elements are optional, and when present, enhance metadata records for NDC users.

### ***REQUIRED METADATA***

Required Element Name	Definition	Number of values
CollectionID <sup>1</sup>	A unique collection ID assigned by the NDC to identify distinct collections. For example, IL field notes collection found here: <a href="https://www.sciencebase.gov/catalog/folder/4f4e49d8e4b07f02db5df141">https://www.sciencebase.gov/catalog/folder/4f4e49d8e4b07f02db5df141</a> , have a CollectionID of 4f4e49d8e4b07f02db5df141. This field may be left blank. NDC staff will assign ID during the file loading process.	1
Title	The human-readable title for the individual record that will be used in any listing or search result. Title should be short for display purposes but contain enough information to distinguish from other records. Examples: Missouri Geological Survey Field Notebooks Missouri Geological Survey Field Notebook Number 428; by Severt, A.F.	1
Abstract	The human-readable description of the individual record used to help determine the nature of the underlying physical data resource. Due to the general nature of the NDC, a fair amount of information about the data resource may need to be captured into this one general element. Example: Sample: Core Research Center, Cutting DD18210, from well operated by St. Michael Exploration, located in Weld County, CO, under lease 2-1 Grace State, with API number 0512310130.	1
DataType	A controlled vocabulary of data types (keywords). An item may include multiple dataTypes, including: 1) Auger Samples, 2) Fluid Samples, 3) Geochemical Samples, 4) Hand Samples, 5) Ice Cores, 6) Paleontological Samples, 7) Rock Cores, 8) Rock Cuttings, 9) Sediment Cores, 10) Sidewall Cores, 11) Thin Sections and Polished	≥1

	Sections, 12) Type Stratigraphic Sections, others.	
SupplementalInformation	This field is used to provide specific information about how to access the physical data represented by the metadata record. This may be general for the entire collection (e.g., a URL to another Web site) or a reference to an online resource such as an ordering system with a specific ID. Example: ResourceID: 2003-01-0050, Repository: Arizona Geological Survey, Contact: Mining Records Curator, Address: 1520 West Adams St Phoenix AZ 85007, E-mail: inquiries@azgs.az.gov, Phone: 602-771-1601	1
Coordinates	Geographic longitude and latitude. Both values shall be contained in the same element and be listed in the order: longitude, latitude with values separated by a comma. Example: Sample: -118.023423, 45.02312	1
DatasetReferenceDate	A reference date indicating currency of the underlying data record, which may be the date the metadata record was assembled for the NDC. Proper date formats are defined in ISO 8601, which include: 1) _yyyy, 2) yyyy-mm, 3) yyyyymmdd, 4) yyyy-mm-dd	1

## **OPTIONAL METADATA**

Required Element Name	Definition	Number of values
AlternateTitle	Collection owners may elect to provide additional title identifiers for individual records for further identification or use by other Web service interfaces. The AlternateTitle field may include either textual titles or specific sample IDs used by the collection.	1
AlternateGeometry	The underlying collection resource may use an alternate method of storing a geospatial footprint. If so, this text field should be used to describe the authoritative source for geographic location and how the simple coordinates were derived.	1
OnlineResource	One or more URL pointers to textual information associated with the record.	≥1
BrowseGraphic	One or more URL pointers to images representing the specific record.	≥1
Date	If a meaningful date within the geosciences domain can be attached to the record (e.g., a collection date), it can be supplied here. Either date may be to any degree of precision, or may be left blank to indicate uncertainty. Examples: 2001, 2001-03, 1939-1945, 20030331, 2000-03-31.	≥1
VerticalExtent	Vertical extent information can be provided and is especially useful for rock core samples. Specification of extent can contain three elements, in following order: unit of measure (meter), maximum depth, minimum depth. Example: m,35.4,0 for a rock core measured at maximum depth of 35.4 meters from the surface.	1

## **2 – Map to National Digital Catalog Metadata**

Having an understanding of the required and optional metadata elements accepted by the NDC, please map the sample level properties found in your own collection to NDC metadata elements. In some cases this might be a one-to-one mapping. For instance, the required “datasetReferenceDate” element might correspond to 'DateCreated' property in your own collection database. For Abstract, it may be desirable to concatenate a number of fields to provide a richer description of the resource and more specific search results. If the location information is stored in township/range/section, it may be necessary to convert these values to a single geographic latitude and longitude point to populate the required coordinates metadata element. However, the descriptive township/range/section may be provided in AlternateGeometry field.

### 3 – Select Upload Format

The National Digital Catalog supports two primary formats for loading data. The simplest format is commonly referred to as a Comma Separated Value or CSV format. This format is fairly easy to create and many software packages use this format to exchange tabular data. The more complex format accepted by NDC is a custom XML format that provides better handling for repeatable elements than possible with the CSV format. The XML format eliminates some parsing issues present in the CSV format. These formats are discussed in more detail below.

#### ***CSV File Format***

A CSV file format represents tabular data. Each row in the file corresponds to a single metadata record, and the properties associated with each record are separated by a comma. The comma is a special character that separates data fields, which is useful for recording record/sample characteristics. However, comma delimited files present issues when data fields include commas, which requires enclosing entire fields in double quotes so that the embedded commas are not regarded as field delimiters but as part of the field information. Further issues arise if data fields include quotes as well as commas, for example, to indicate inches in a measurement (one inch represented as 1”). To avoid these issues, NDC requests data providers use “pipe” character (|) to delimit fields rather than a comma. The pipe character is seldom used in normal text and allows fields to include the more common comma and quote characters without risking confusion with the field separator. For example, a single record with four fields might look like the following using the pipe record delimiter character:

```
1341234|This is a Title|m,35.4,0|Please note, the existence of a comma and “quote” characters.
```

The above line could not be represented accurately using a comma delimited approach, but is fairly straightforward using the pipe delimiter character.

CSV formatted files must include a header on the first line that indicates the metadata element names and delimiters corresponding to the subsequent data (similar to a map legend). Applying this to the above example, the first two lines of the CSV file would look like the following:

```
COLLECTIONID|TITLE|VERTICALEXTENT|ABSTRACT  
1341234|This is a Title|m,35.4,0|Please note, the existence of commas and quote characters.
```

Most databases may output CSV formatted files through built-in utilities, or customized queries. Microsoft Excel may be used to produce

CSV formatted files but the options for character encoding and delimiters are limited. Excel uses the system settings to determine character encoding and delimiters (comma versus pipe). On a Windows platform, delimiters may be set by accessing the Control Panel --> Regional and Language Options. This option is not available on a Macintosh.

## ***XML File Format***

The XML file format is more complex than the CSV, easier for describing multiple values, and is more reliable for handling various data inputs. It may be helpful to use XML formatting tools, available online, when working with XML format. Some databases may produce information in an XML format, which may be not be acceptable NDC but may be transformed using an additional XSL stylesheet into the correct format.

An online XML template is available at: <http://datapreservation.usgs.gov/docs/collectionMetadataExample.xml>

A basic two-record sample is available at: <http://datapreservation.usgs.gov/docs/NGGDPPSampleMetadata1.xml>

XML schema is provided for basic XML validation: <http://datapreservation.usgs.gov/docs/NGGDPPMetadata.xsd>

## **4 – Upload Metadata to National Digital Catalog**

To view geoscience collections and samples provided by your state to the NDC, visit the NDC Dashboard at [www.sciencebase.gov/ndcDashboard](http://www.sciencebase.gov/ndcDashboard)

## **Contact Information**

Natalie Latysh, Associate National Geological and Geophysical Data Preservation Program Coordinator  
303-202-4852  
[nlatysh@usgs.gov](mailto:nlatysh@usgs.gov)

Betty M. Adrian, National Geological and Geophysical Data Preservation Program Coordinator  
U.S. Geological Survey  
303-202-4828  
[badrian@usgs.gov](mailto:badrian@usgs.gov)